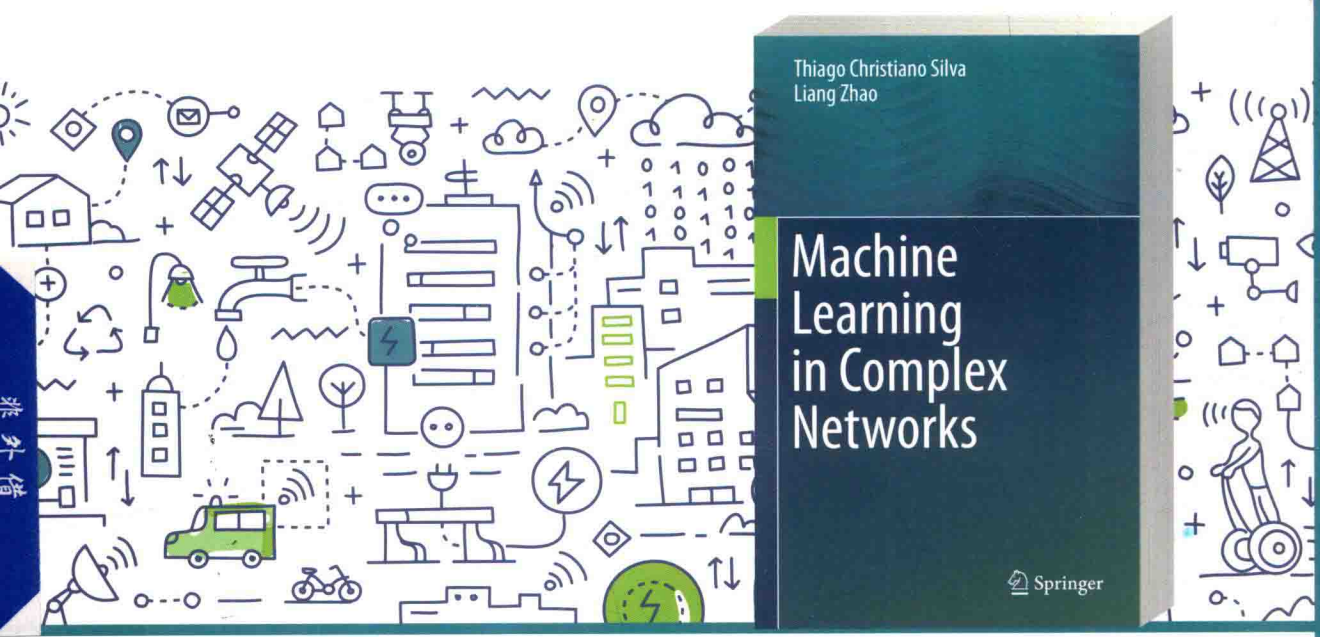
HZ BOOKS 华章科技
Springer
智能科学与技术丛书

Machine Learning in Complex Networks

基于复杂网络的机器学习方法

[巴西] 迪亚戈·克里斯蒂亚诺·席尔瓦（Thiago Christiano Silva）◎著
赵亮（Liang Zhao）
李泽荃 杨曌 陈欣 ◎译



机械工业出版社
China Machine Press

智能科学与技术丛书

Machine Learning in Complex Networks

基于复杂网络的机器学习方法

[巴西] 迪亚戈·克里斯蒂亚诺·席尔瓦（Thiago Christiano Silva）
赵亮（Liang Zhao） ◎著

李泽荃 杨曌 陈欣 ◎译

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

基于复杂网络的机器学习方法 /（巴西）迪亚戈·克里斯蒂亚诺·席尔瓦（Thiago Christiano Silva），赵亮（Liang Zhao）著；李泽荃，杨曌，陈欣译．—北京：机械工业出版社，2018.10
（智能科学与技术丛书）
书名原文：Machine Learning in Complex Networks

ISBN 978-7-111-61149-3

I. 基… II. ①迪… ②赵… ③李… ④杨… ⑤陈… III. 机器学习 IV. TP181

中国版本图书馆 CIP 数据核字（2018）第 235238 号

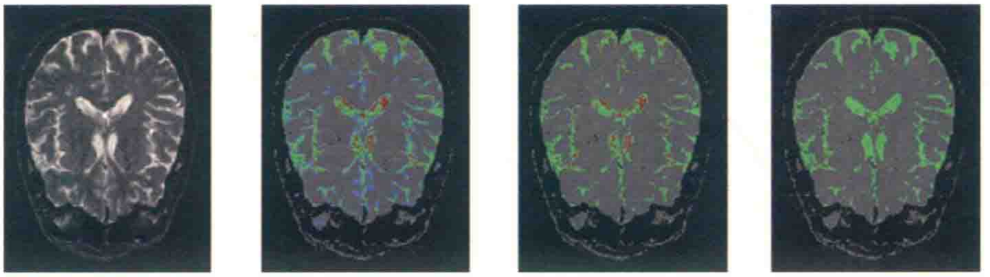
本书版权登记号：图字 01-2018-1370

本书将机器学习和复杂网络这两个重要的研究方向结合起来，不仅包括必备的基础知识，还涵盖新近的研究成果。书中首先介绍机器学习和复杂网络的基本概念，然后描述基于网络的机器学习技术，最后对监督学习、无监督学习和半监督学习方法的案例进行详细分析。

本书通过大量例子和图示来帮助读者理解各类方法的主要思路和实现细节，并列出了可供深入研究的参考文献，适合该领域的研究人员、技术人员和学生阅读参考。

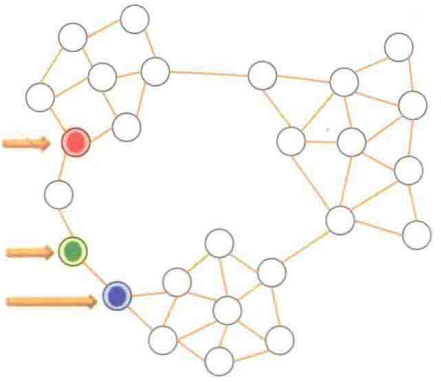
出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）
责任编辑：曲 熠　　责任校对：殷 虹
印　　刷：三河市宏图印务有限公司　　版　　次：2018 年 11 月第 1 版第 1 次印刷
开　　本：185mm × 260mm 1/16　　印　　张：16.5（含 0.25 印张彩插）
书　　号：ISBN 978-7-111-61149-3　　定　　价：79.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：(010) 88378991 88361066　　投稿热线：(010) 88379604
购书热线：(010) 68326294 88379649 68995259　　读者信箱：hzjsj@hzbook.com

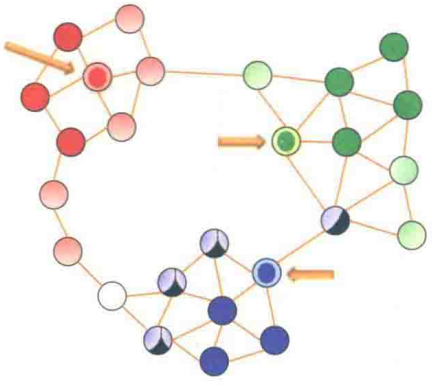


a) 大脑图像　　b) 五个社团　　c) 四个社团　　d) 三个社团

图 4-10　人类大脑的像素聚类。图中的颜色表示聚类，$d=0.008$，$k=0.5d$



a) 可能的初始状态



b) 预期的长时动力学过程

图 9-1　粒子竞争模型的初始条件和长时动力学过程

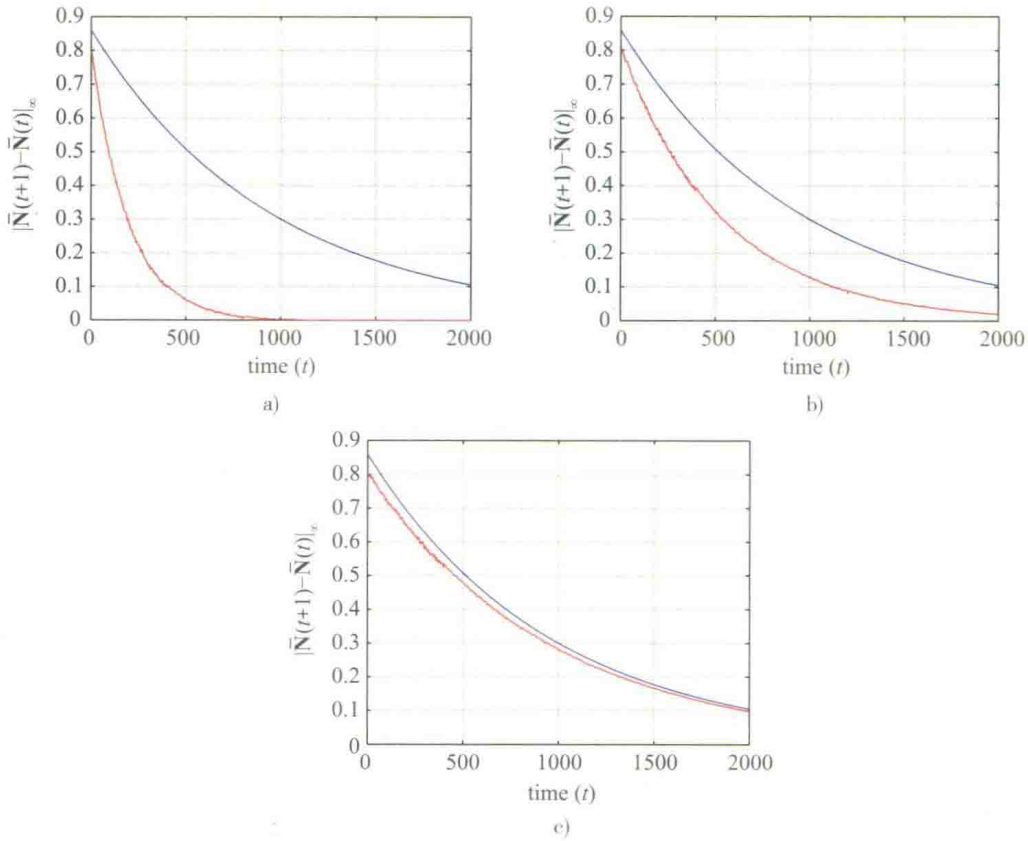


图 9-9　考虑 $|\bar{\mathbf{N}}(t+1)-\bar{\mathbf{N}}(t)|_\infty$ 的粒子竞争模型收敛分析[36]。模拟基于图 9-7 中的数据集：当 $c=K$ 时，理论解的上限用蓝色曲线表示

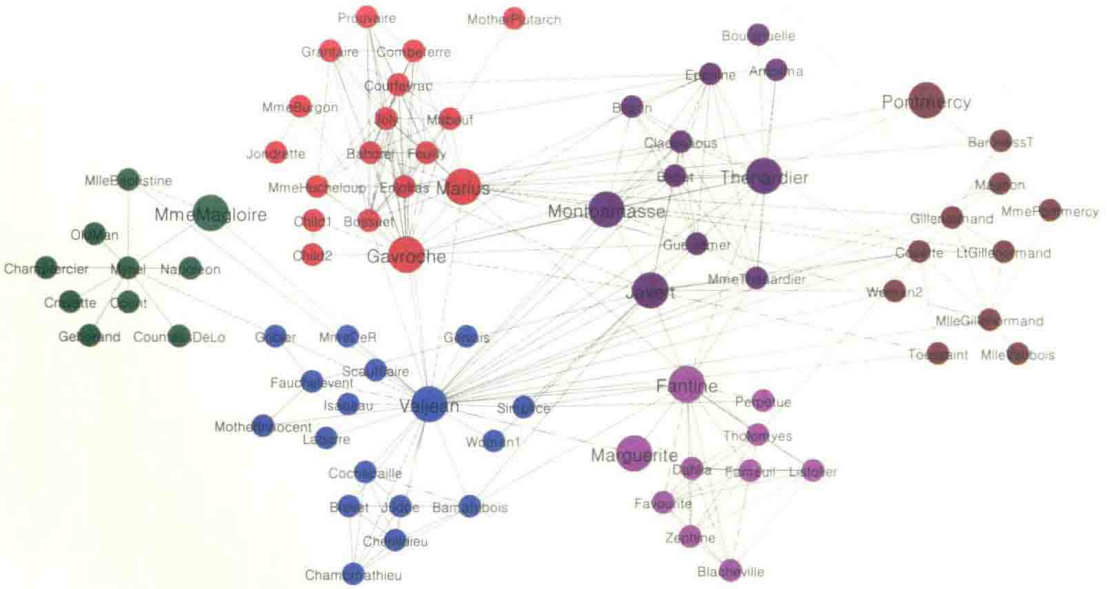


图 9-14　维克多·雨果的作品《悲惨世界》中主要人物的关系网络。$K=6$，$\lambda=0.6$；10 个最大重叠度的节点以较大尺寸显示

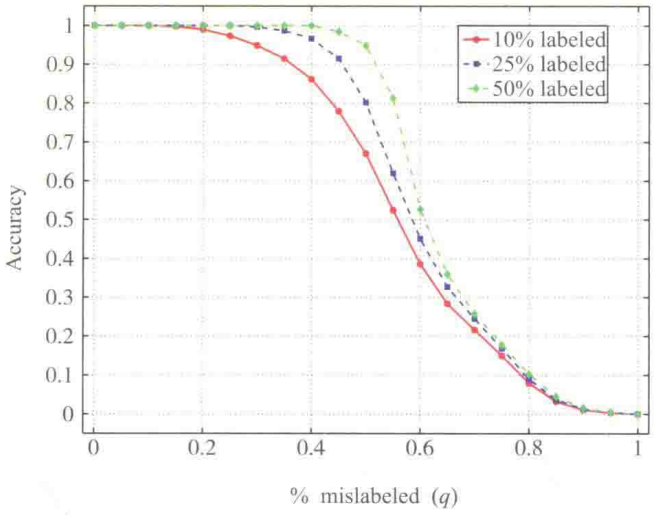
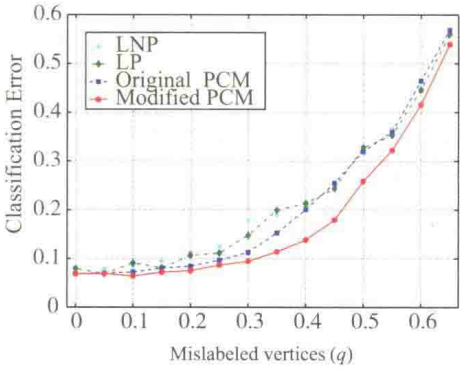
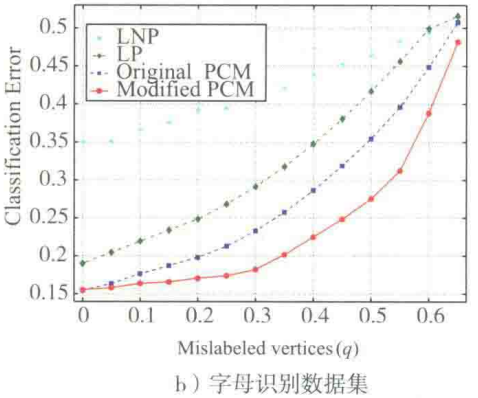


图 10-9　模型的准确率与无标签或错误标记样本比例 q 的关系。网络中社团类别恒定，共 16 个社团；节点数为 10000；社团重叠程度 $\frac{z_{out}}{\langle k\rangle}=0.3$；仿真 100 次取平均值



a）鸢尾花数据集



b）字母识别数据集

图 10-10　两个真实数据集上噪声数据比例的测试。运行 100 次并取平均值。经 Elsevier 许可从文献［12］转载

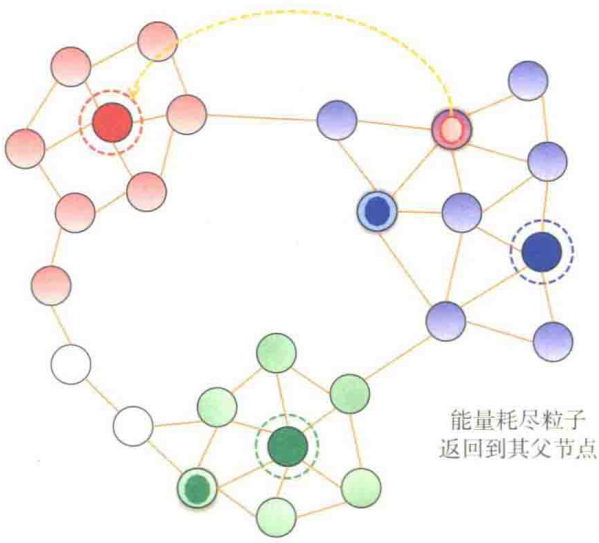
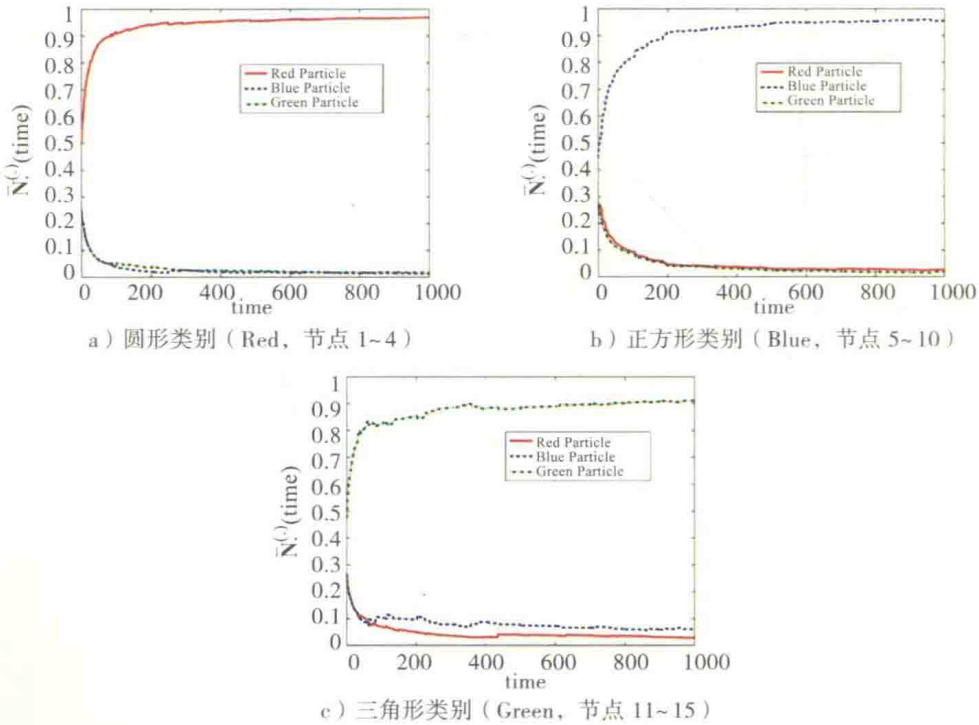


图 10-3　改进后的能量耗尽粒子复苏过程。带虚线圆圈的节点表示已标记节点，节点的颜色深度表示拥有最高控制能力粒子的颜色



a）圆形类别（Red，节点 1~4）　　b）正方形类别（Blue，节点 5~10）

c）三角形类别（Green，节点 11~15）

图 10-5　网络中三个粒子平均控制能力的演化过程

2017 年年初，谷歌旗下 DeepMind 团队开发的 AlphaGo 升级版 Master 战胜了柯洁、陈耀烨、李世石、三村智保等中日韩顶尖围棋手，取得 60 胜 0 负的辉煌战绩。这又一次促使人工智能相关话题迅速升温，越来越多的公司开始开展人工智能技术研究，越来越多的从业人员开始进入该领域寻求机会。同时，人工智能技术确实也在影响着我们的生活，如帮助医生进行医疗诊断，帮助房产公司评估资产价值，帮助物流公司规划路径等。

机器学习是实现人工智能的一种方法，其概念来自早期的人工智能学者。机器学习分为监督学习、半监督学习和无监督学习，目前常用的算法有决策树、逻辑斯谛回归、朴素贝叶斯、k-均值等。简单来说，机器学习就是使用算法分析数据，并根据学习到的模型做出推断或预测。

虽然已经有众多的机器学习方法被提出并且在各类实际系统中成功应用，但是仍然有很多挑战性的问题需要解决。近年来，随着社交网络的快速发展，数据规模暴增，特别是本身就呈现网络特征的数据样本急剧增加，促使基于复杂网络的机器学习方法被广泛关注。

本书的两位作者 Thiago Christiano Silva 和 Liang Zhao 长期从事复杂网络和机器学习的交叉研究。他们深知基于网络的机器学习技术的内在优点，并经多方面调研论证，将多年的研究成果汇聚成书，供各领域的研究人员参考学习。

这种学科交叉融合带来的良性互动，无疑促进了包括复杂网络、机器学习在内的诸多学科的繁荣。这也正是本书的目的和意义。

感谢本书的作者 Liang Zhao 教授给予大力支持，他提供的方便使得本书的翻译工作能够及时完成。

感谢机械工业出版社华章公司的编辑，是他们的远见使得本书能够快速与读者见面。

感谢第一译者的爱人屯超参与我们的校对工作。

由于译者水平有限，译文中难免出现词不达意的问题。文中的错误和不当之处，希望读者与我们联系，以便不断改进。意见请发往 lzquancumtb@126.com 或 yangzhaocumtb@126.com，我们将不胜感激。

李泽荃

2018 年 6 月 1 日于北京

前　言

Machine Learning in Complex Networks

机器学习是计算机科学的一个重要研究领域之一，主要指计算机利用已有的经验来获得学习能力的一种计算方法。虽然已经有众多的机器学习方法被提出并且在各类实际系统中成功应用，但是仍然有很多挑战性的问题需要解决。在过去的几年里，基于复杂网络（大规模的具有复杂连接模式的图）的机器学习方法越来越受到关注。该方法的出现是因为其具有内在的优点，即数据表示是基于网络特性的，能有效捕获数据的空间、拓扑和功能关系。本书介绍了在机器学习领域复杂网络理论的特性和优势。在前七章，我们首先介绍机器学习和复杂网络的一些基本概念，提供必要的背景知识。然后，简要描述基于网络的机器学习技术。在后三章，我们将介绍一些基于网络的监督学习、无监督学习和半监督学习方法，并提供详细的案例分析。特别是，针对无监督和半监督学习，我们探讨了使用随机非线性动力系统的粒子竞争技术。同时，分析了竞争系统内的各类影响因素，以确保该技术的有效性。另外，对于学习系统存在的不完善性，比如半监督学习的数据可靠性问题，可以采用竞争机制来消除训练数据集的缺陷。识别并预防误差传播具有重要的实际意义，但文献中关于这方面的研究很少。在案例分析中，我们提出了一个结合低阶和高阶的混合监督分类技术，低阶项通过传统的分类方法实现，而高阶项通过提取由输入数据构造的底层网络的特征实现。换句话说，其主要思路是低阶项利用数据的物理特征实现测试样本的分类，而高阶项进行测试样本模式的一致性检验。可以看出，该技术可以根据数据的语义特征实现样本分类。

本书旨在融合两个目前被广泛研究的领域：机器学习和复杂网络。所以，我们希望本书能在科学界引起更多学者的兴趣。本书是自成体系的，介绍基于网络的机器学习技术的建模、分析和应用，不仅包含两个领域的基础知识，还介绍了一些新的研究成果，主要面向对机器学习和复杂网络感兴趣的研究人员和学生。对于每一个可探索的话题，我们还提供了相应的参考文献。此外，众多的说明性图例也可以帮助读者理解各类方法的主要思路和实现细节。

致谢

感谢 Marcos Gonçalves Quiles 博士、Fabricio Aparecido Breve 博士、João Roberto Bertini Jr. 博士、Thiago Henrique Cupertino 博士、Andrés Eduardo Coca Salazar 博士、Bilzã Marques de Araújo 博士、Thiago Ferreira Covões 博士、Elbert Einstein Nehrer

Macau 博士、Alneu Andrade Lopes 博士、Xiaoming Liang 博士、Zonghua Liu 博士、Antonio Paulo Galdeano Damiance Junior 先生、Tatyana Bitencourt Soares de Oliveira 女士、Lilian Berton 女士、Jean Pierre Huertas Lopez 先生、Murillo Guimarães Carneiro 先生、Leonardo Nascimento Ferreira 先生、Fabio Willian Zamoner 先生、Roberto Alves Gueleri 先生、Fabiano Berardo de Sousa 先生、Filipe Alves Neto Verri 先生和 Paulo Roberto Urio 先生过去的几年里在该领域内的合作。感谢 Jorge Nakahara Jr. 博士仔细审阅了本书，并在整个出版过程中给予我们持续支持。感谢 Ying-Cheng Lai 博士引导我们进入迷人的复杂网络研究领域。感谢 Hamlet Pessoa Farias Junior 先生和 Victor Dolirio Ferreira Barbosa 先生热烈的讨论成果。也要感谢 João Eliakin Mota de Oliveira 先生为我们提供了两张图。同时，感谢巴西圣保罗大学数学与计算机科学研究所（ICMC）和里贝朗普雷图分校哲学、科学与文学学院，以及巴西中央银行的大力支持。最后，感谢巴西圣保罗研究基金会（FAPESP）、巴西国家科学技术发展委员会（CNPq）和巴西高等教育基金会（CAPES）为我们的研究工作提供资金支持。

Thiago Christiano Silva

Liang Zhao

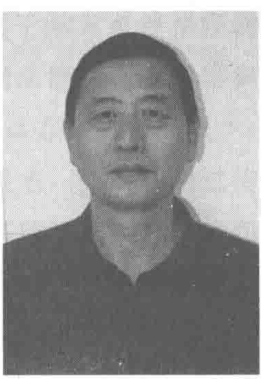
巴西，巴西利亚和里贝朗普雷图

2015 年 11 月

作者简介

Machine Learning in Complex Networks

Thiago Christiano Silva 博士于 2009 年在巴西圣保罗大学以第一名的成绩获得计算机工程学士学位。2012 年在圣保罗大学获得数学与计算机科学博士学位，并由于其开创性的博士论文而获得众多奖项。2013 年，他是“计算机科学领域的 CAPES 论文竞赛”“圣保罗大学论文竞赛”“国际 BRICS-CCI 博士论文竞赛”的获胜者。2014 年，他在圣保罗大学完成了一年的机器学习和复杂网络博士后研究。自 2011 年以来，他一直在巴西中央银行担任研究员。他的研究领域包括机器学习、复杂网络、金融稳定性、系统风险和银行业务等。



Liang Zhao 博士于 1988 年在武汉大学获得计算机科学学士学位，于 1996 年和 1998 年在巴西航空技术学院分别获得计算机科学硕士和博士学位。他在 2000 年以教师身份加入了圣保罗大学的数学与计算机科学学院。目前，他是圣保罗大学哲学、科学与文学学院计算机科学与数学系的系主任、教授。2003 至 2004 年，他是美国亚利桑那州立大学数学系的客座研究员。他目前的研究兴趣包括机器学习、复杂网络、人工神经网络和模式识别。他在国际期刊、书籍和会议上发表了超过 180 篇学术论文。他是巴西优秀科研人员津贴获得者。他目前是《Neural Networks》的副主编，并于 2009 至 2012 年担任《IEEE Transactions on Neural Networks and Learning Systems》的副主编。他是 IEEE 高级会员，也是国际神经网络协会（INNS）的成员。

$\mathcal{G}$	图或网络
$\mathcal{V}$	节点集
$\mathcal{E}$	连边集
$\mathcal{L}$	已标记训练数据集
$\mathcal{U}$	未标记数据集
$\mathcal{Y}$	标签或目标集
$\mathcal{K}$	粒子集
$\mathcal{N}(v)$	节点 v 的邻居节点集
$\mathcal{X}$	向量形式的数据样本集
$\mathcal{X}_L$	向量形式的已标记数据样本集
$\mathcal{X}_U$	向量形式的未标记数据样本集
$\mathcal{X}_{\text{training}}$	训练样本集
$\mathcal{X}_{\text{test}}$	测试样本集
$\mathcal{O}(.)$	算法的时间复杂性
V	节点数量
E	连边数量
L	已标记训练样本数量
U	未标记样本数量
Y	标签或目标数量
K	粒子数量
P	样本的特征（维度）数量
M	样本集中的社团数量
λ	随机规则和优先规则之间的平衡因子
ρ	依从项
$\mathbf{A}$	加权或非加权邻接矩阵
$\mathbf{P}$	转移矩阵
$\mathbf{R}$	势矩阵
$\mathbf{S}$	相似矩阵
$\mathbf{D}$	相异矩阵
k_v	节点 v 的度
$k_v^{(\text{in})}$	节点 v 的入度

$k_v^{(\text{out})}$	节点 v 的出度
s_v	节点 v 的强度
$s_v^{(\text{in})}$	节点 v 的入强度
$s_v^{(\text{out})}$	节点 v 的出强度
CC_v	节点 v 的聚类系数
$\overline{k}$	平均度或平均连通度
r	网络同配性
Q	网络模块化
CC	网络平均聚类系数
d_{uv}	节点 u 到 v 的最短路径长度

目　录

Machine Learning in Complex Networks